大数据时代：地质学的挑战与机遇

翟明国[1,2*] 杨树锋[1*] 陈宁华[1] 陈汉林[1]

1 浙江大学 地球科学学院 杭州 310027

2 中国科学院地质与地球物理研究所 北京 100012

摘要 大数据正在影响着人类生活，改变着人类认识和研究世界的思维方式。作为典型数据密集型学科的地质学，正面临着前所未有的挑战与机遇。为了应对这一挑战，地质学家不仅需要改进传统的研究方法，更重要的是要改变传统思维模式，拥抱大数据时代的到来。地质学与大数据的结合不仅极大拓展了地质学的认知空间，提升了获取地质学新知识的能力，同时为地质学支撑的能源矿产调查、环境资源合理利用以及防灾减灾等社会生产和公共服务提供了创新活力。在分析国内地质学大数据研究现状的基础上，文章阐述了我国地质学大数据研究的前沿科学问题，提出了地质学大数据发展战略目标，探讨了地质学大数据发展面临的主要问题和解决途径。大数据将改变地质学家的思维方式，数据驱动的科学发现模式将为地质学的发展带来全新的面貌，文章呼吁地质学界对大数据给予更多的关注和支持。

关键词 地质学，大数据，数据密集型，数据挖掘

DOI 10.16418/j.issn.1000-3045.2018.08.009

地质学是研究地球如何演化的自然科学，以固体岩石圈为主要研究对象，探讨地球各圈层的物质组成、内部构造、外部特征、各层圈之间的相互作用和演变历史。地质学的产生源于人类社会对石油、煤炭、金属、非金属等矿产资源的需求，随着社会生产力的发展，人类活动对地球的影响越来越大，地质环境对人类的制约作用也越来越明显。如何合理有效地利用地球资源、维护人类生存的环境，已成为当今世界所共同关注的问题。因此，地质学研究领域进一步拓展到研究地球资源的合理利用、资源与环境和人-地和谐的可持续性发展。

地质学是通过对自然现象的观察，发现观测数据中内在的规律性，其本质上是一门信息学科，是典型的数据密集型科学。21 世纪以来，随着地球信息探测技术的日新月异，获取数据的能力不断提高，积累的地球观测数据呈指数级增长，预计到 2020 年全球数据总量将达到 40 ZB[1]。多元、多维、多源、异构、时空性、方向性、相关性、随机性、模糊性、时空不均匀性和过程的非线性是地质大数据的特点。地质大数据与一般大数据

* 通讯作者

资助项目：中国科学院学部咨询项目（81721360）

修改稿收到日期：2018年8月13日

有相似之处，但也存在显著差别，大数据给地质学带来了前所未有的机遇与挑战。

当前，大数据正在影响着人类生活，改变着人类认识和研究世界的思维方式[2]。大数据时代，数据密集型知识发现成为继理论科学、实验科学和计算科学后科学研究的第四范式。2008 年、2011 年 *Nature* 和 *Science* 分别出版了大数据研究的专刊。大数据研究成为各国关注和优先发展的国家战略性技术。作为国家大数据战略的重要组成部分，地质大数据的应用研究方兴未艾。《国土资源“十三五”科技创新发展规划》指出，要强化地学基础研究，开展地球深部过程与动力学、地球环境演化与生命过程、矿产资源和化石能源形成机理研究，加强地球关键带过程与功能、全球环境变化与地球圈层相互作用、人类活动对环境影响、重大灾害形成机理研究，深化地学大数据与地球系统知识发现研究。云计算、物联网、人工智能等技术的兴起，使信息技术渗透方式与处理方法及应用模式发生变革、地质研究中多系统联合与结合成为可能[3]。

利用地质学与大数据的结合，开展对地质数据的综合研究和应用，拓展了地质学的认知空间，提升了获取地质学新知识的能力。相比大气科学、海洋科学、遥感科学以及全球变化等地学大数据的研究，地质学大数据的研究起步较晚，由于缺乏顶层设计和基本规划，地质大数据中心建设薄弱。本文着眼于我国未来地质学发展，在分析地质大数据特点及国内地质大数据研究现状的基础上，阐述地质大数据拟研究的前沿科学问题，提出了未来中国地质大数据发展战略目标，探讨了地质学大数据发展的主要内容和解决途径。

1 地质大数据的特点

地质大数据是通过露头地质观测、勘查工程、地球物理探测、地球化学探测、遥感和物理测试、化学分析等手段采集到的一种科学大数据[4]，涉及地球从内到外的各个圈层，涉及地球形成与演化的历史，涉及地球的物质组成及其变化，涉及矿产资源的形成、勘查与开发利用，涉及人类环境的破坏与修复等[5]。

地质大数据具有传统大数据的“4V”特性，即数据量大（volume）、类型繁多（variety）、速度快时效高（velocity）、价值密度低（value），同时还具有科学大数据的“三高”特点，即高维度（high dimension）、高计算复杂性（high complexity）和高不确定性（high uncertainty），且由于地质对象的发展演化时空范围庞大、地质作用影响因素众多，这种高维度、高计算复杂性和高不确定性特点则更加显著。地质大数据的特点主要表现在以下 4 个方面。

1.1 多源（元）异构性

地质数据的采集平台和手段繁多，不同手段获取的数据也具有不同的数据组织管理形式。例如：野外露头描述数据、钻孔岩芯描述数据、各种地质报告文档数据，以及大量野外填图、素描和照片，遥感获取的影像数据，地质灾害监测获得实时点位数据等。有的数据是以纸质形式存储和管理，有的经过结构化转换汇聚 GIS 并入库存储。不同的数据组织方式形成了不同的数据结构，对同一地质本体的描述也因空间基准和时空尺度存在差异而形成语义鸿沟。不同的数据采集方法，多角度的描述，造成了地质大数据的严重异构和多模态。

1.2 时空相关性

由于地质体、地质结构、地质资源、地质环境和地质灾害通常占据庞大的空间范围且伴随长时序的发展演化，地质大数据的“时间维”具有长期性和阶段性的复杂特点，是其他地球科学数据不具备的。

地质数据高度的时空特性体现在两方面：① 地质对象本身具有特定的地质年代，地质学研究也具有明显的区域性，在特定时段、特定区域内的研究对象往往带有明显的差异化特征。② 地质数据用来描述某一时间点在特定位置的对象的属性，这种固有属性在获取数据时就已经产生。地质数据的时间尺度可从分秒横跨至数十万

年，再加上描述其空间位置的坐标系、投影参数以及探测精度、基本粒度的变化，则进一步增加了地质数据的复杂度。因此，脱离了时间、空间的地质数据是没有意义的，在涉及地质大数据的融合分析时也要将其划归到统一的时空基准下。

1.3 复杂性与模糊性

地球是一个复杂的巨系统，地质数据的参与在一定程度上降低了该系统的复杂度，使建模和求解成为可能。然而由于地球各圈层因素相互作用，各类地质过程本身具有高度的复杂性，人类对于许多地质规律的解释和结论还存在争议。再加上地质数据对描述对象的定量化困难，决定了地质数据分析、建模和计算的困难程度。大数据的技术取向之一是“重关联不重因果”，我们不可能仅通过数据就搞清楚地质现象发生的机理，且采集全球样本数据到现阶段还无法实现，因此地质大数据分析的结果大多是模糊不确定的[6]。

1.4 地质体的全球性与国家利益

地质体和地质单元的分布不以国家界限为界限，地质资源的分布不以国家和人口的需求而分布。这就造成“国家利益”干预造成的全球数据库建设的困难。

2 地质大数据研究进展

总体来看，我国地质大数据研究还处于起步阶段。一方面，许多人质疑大数据对于地球科学这一门观察学科的适用性，另一方面，多数研究人员还未意识到积累数据、共享数据的重要性，这在一定程度上阻碍了地质大数据的发展[7]。此外，大数据研究追求的“相关性”研究与科学研究中的“因果性”知识发现存在巨大矛盾，这对科学家们的思维方式转变提出了挑战[8]。综合近年来的研究成果，国内地质大数据研究现状可以归结为以下3点。

2.1 地质大数据的存储管理

地质学发展至今积累了大量的地质资料数据，随着地球信息探测技术的迅速发展，又有源源不断的新的地质数据快速产生。地质大数据不仅有定性、定量数据，还包括文字说明，甚至是地质图件或者是地质工作者在工作中留下的视频、音频文件等资料，而长期的目录文件存储方式极大地降低了数据查询、检索、统计、更新、挖掘等操作效率，导致数据服务能力低下[3]。因此，构建一套能够有效地实现结构化、半结构化和非结构化数据一体化、静态数据与动态数据一体化、地质数据与地质模型一体化存储管理的地质信息系统，对于完成海量地质资料稳定、高效地存放与读取就显得十分重要[4]。

目前已有学者提出进利用云平台、Hadoop 和 NoSQL 等技术，借鉴实时 GIS 时空数据模型[9]，实现对地质时空大数据模型的动态管理。Hadoop 是目前大数据存储与处理的标准平台，可以通过 MapReduce 支持大规模数据的并行处理。而 NoSQL 数据库使用分布式节点集动态处理负载。采用分布式文件系统技术可以对地质大数据进行存储并提高数据的容错能力与可靠性[10]。例如中国地质调查局发展研究中心国土资源部地质信息技术重点实验室研发的中国地质调查云平台，就是在这种框架下建立的非结构化地质数据存储组织模式，通过改变非结构化数据的存储、阅读、搜索和应用模式，为智能地质调查提供精确、快速服务奠定了基础[11]。

2.2 地质大数据的挖掘分析

大数据时代下的 3 个重要技术取向是：要全体不要抽样；要效率不要绝对精确；要关联不要因果[7]。这迫使我们从数据的类型、数据运维以及大数据带来的挑战性这 3 个维度重新思考数据分析。周永章等[12]认为大数据与数学地球科学的核心应用技术应该包括高维数据降维、图像数据处理、无限数据流挖掘、机器学习、关联规则算法与推荐系统算法等。

数据挖掘是指从大量数据中通过算法搜索其隐藏信息的过程[13]。相较于数据检索和信息提取，数据挖掘需要基于大数据和知识库的智能推理的理论和技术支撑[14]。地质大数据挖掘就是从数据仓库中找寻隐含的特征和规律，并应用在地质规律研究、成矿预测、资源评价、环

境保护和地灾防治领域的过程。该过程需要利用涉及到人工智能、机器学习、模式识别、归纳推理、统计学、数据库、高性能计算、数据可视化等相关方法和技术手段，在多主题、多模态的地质数据中自动或半自动地获取新的可被理解的知识，从而为地质专题研究和应用提供决策。

目前，数字地质的任务就是大力推动地质科学的数据挖掘和数据分析方法的更新。如何从规模巨量，但价值密度偏低的大数据中有效地挖掘提取信息是当下地质大数据研究中拟解决的关键问题。地质大数据分析的关键技术主要是对多源（元）异构的地质数据进行综合分析。其中包括对结构化数据的相关性分析，对半结构化数据的信息提取和结合非结构化数据作为以上数据处理结果的验证分析。此外，物联网、虚拟现实、云计算等技术兴起，使得基于互联网的地质数据资源共享平台的研制成为可能，也为复杂地学计算提供了条件。将云计算、人工智能融入地质大数据挖掘与分析已经成为新的发展趋势。例如，有学者借鉴了大数据思维，利用贝叶斯网络探寻矿床的成因机制，从而构建大数据-智能矿床成矿与找矿模型[15]，推动从“数字地质”到“智能地质”的革命。

2.3 地质大数据的应用服务

地质大数据不仅改变了地质学家研究科学问题的思维范式，也给以数据分析为基础的地质行业带来了技术革新。地质大数据在各领域数据化水平的提高，有效地打通了信息孤岛，使定量化分析能够进一步推进。地质大数据的应用服务主要体现在以下5个方面。

（1）基础地质调查。《国土资源“十三五”科技创新发展规划》指出要推进数字地质调查系统向智能化方向发展，逐步实现地质数据快速采集、实时汇聚、高效分析处理与建模，推动大数据技术支撑下的智能地质调查和服务模式创新，深化地质填图、矿产地质调查、油气地质调查、海岸带综合地质调查等领域的应用。

如何将分布式的数据云存储、云管理和云服务体系应用在我国各类基础地质调查数据库，实现海量、碎片化、非结构化与多样性的数据高效快速存储，是大数据时代基础地质调查研究的热点[16]。此外，我国正在开展数字地质调查，中国地质调查局开发的“地质云1.0”已经在2017年正式发布并上线服务。该系统面向各类地质调查专业人员提供基础地质、矿产地质、水工环地质、海洋地质等多类专业数据共享服务；面向社会公众提供多类地质信息产品服务。升级完善的智能地质调查系统已在基础地质和矿产地质调查领域示范应用。

（2）国土资源管理。国土资源部门在多年的信息化建设实践中积累了海量的土地数据，进而提出了国土资源全尺度数据整合与大数据构建技术。2016年国土资源部提出要持续完善国土资源“一张图”数据资源体系，构建统一的国土资源数据共享和开放平台。其中大数据采集与分析技术成为构建决策支持系统、智库信息化工作平台，逐步形成信息化条件下的新型“互联网+”智库运行体系的重要技术手段，对于提升国土资源宏观调控、管理监测、形势分析、政策评估、舆情分析等领域具有重要的决策支持能力。

（3）地质灾害监测。以物联网、大数据技术为支撑，从海量地质灾害数据中充分挖掘数据的潜在信息价值，并结合多轨道、多尺度和多时相的遥感环境监测技术，建立智能化的地质灾害、地下水、矿山地质环境、地面沉降、水土环境、地质遗迹等调查、监测数据采集系统和预警预报系统，从而加强对灾害发生趋势的研判和预测，强化实时监测与预警，用数据的力量防治地质灾害。

（4）矿产资源勘查。矿产资源是国民经济发展所需的重要物质基础，而矿产资源预测是资源发现与勘察中的指导性工作。以往专业人员都是在一定的理论和方法指导下，凭借已有的知识和经验并采用定性或定量的方法进行预测找矿。而随着矿产资源预测理论的不断进步，以及地学信息与虚拟现实技术、3S技术、数据库技术、三维建模及可视化技术等的有机融合，对于认识新

的成矿规律意义重大。这种方法从地质科学相关的海量数据中进行挖掘，对各种矿床类型进行多维度、多特征的描述和建模，从而代替由少量参数构成的预测模型，实现了地质理论和实际问题解决、数学应用和数学模型研究与信息技术应用三结合的矿产资源预测评价[17]。此外，大数据驱动下的成矿预测理论的出现，进一步催生了大量以空间数据库为基础的三维可视化软件系统和矿产资源预测系统，为智慧找矿奠定了基础[18]。

（5）三维可视化。数据可视化是描述、表达和理解各种半结构化甚至非结构化问题的关系和模型的最佳方法和手段[19]。以地质空间大数据为基础，结合三维可视化、虚拟现实技术等，针对地质体和地质结构进行三维动态可视化建模，则可构成“玻璃地球”，帮助科研人员分析、预测、评估和决策。以数字矿山技术发展为例，三维可视化技术能够更加生动地展示矿山地质地貌的信息，清楚地反映矿体赋存状态[20]，从而综合、动态地指导研究人员进行矿体定位与成矿预测工作。

3 结语

大数据时代给地质学的发展带来了机遇和挑战。一方面，地质大数据为我们全面感知、了解地球打开了新的图景，也为地质科学的知识发现、科技创新提供了新的手段和途径。另一方面，由于地质大数据具有科学大数据的“三高”特点，给大数据的挖掘和利用提出了难题。此外，数据交流、共享机制的不成熟也成为地质大数据研究发展的阻碍之一。如何建立高效的大数据服务平台，推动具有大数据源的各个学科协同研究是今后要解决的重要问题。我国地质大数据的研究还处于起步阶段，但它的重要战略意义和发展应用前景都应该得到肯定。为此，提出加快我国地质大数据建设进程的3点建议。

（1）推动“地质+大数据”人才培养体系的建立。高校要应对大数据时代的挑战，建立地质大数据人才培养方案。呼吁教育部和科技部加大对地质大数据项目的支持，以项目育人才，培育出既有扎实的地质学基础，同时熟悉算法开发、数据建模及数据架构，并且能够胜任地质大数据系统研发、地质大数据挖掘与分析、地质大数据应用开发等工作的学科交叉型专业人才。

（2）加快地质大数据共享交流平台的建立。数据的自由流通和共建共享是发挥数据资源价值的关键。目前大多地质数据资源建设都由重大科研项目驱动，有一定的项目实施周期，其数据服务平台也存在着功能单一、检索效率低下、数据库建设标准不一等缺点，造成系统平台数据流通性、可用性较差。应由国家层面的专门机构协调构建由高校、科研院所和地质生产单位共同参与的地质大数据中心，在保护国家利益的前提下，加快构建规范统一的地质大数据共享交流平台，推动地质大数据的研究与应用。

（3）地质学家和地质工作人员的思维变革。科学大数据已成为科学研究的重要途径，数据密集型科学范式也已逐渐被接受。地质学家和广大地质工作人员应该抓住历史契机，拥抱大数据，改变经验的传统思维模式，以新的态度看待数据，以新的思维方式利用数据，从中获取新知识，创造新价值。

参考文献

1 Gantz J, Reinsel D. The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. Framingham: IDC Analyze the Future, 2012.

2 赵鹏大. 大数据时代呼唤各科学领域的数据科学. 中国科技奖励, 2014, (9): 29-30.

3 陈建平, 李靖, 谢帅, 等. 中国地质大数据研究现状. 地质学刊, 2017, 41(3): 353-366.

4 吴冲龙, 刘刚, 张夏林, 等. 地质科学大数据及其利用的若干问题探讨. 科学通报, 2016, 61(16): 1797-1807.

5 王登红, 刘新星, 刘丽君. 地质大数据的特点及其在成矿规律、成矿系列研究中的应用. 矿床地质, 2015, 34(6): 1143-1154.

6 谭永杰, 文敏, 朱月琴, 等. 地质数据的大数据特性研究. 中国

矿业, 2017, 26(9): 67-71.

7 张旗, 周永章. 大数据正在引发地球科学领域一场深刻的革命——《地质科学》2017年大数据专题代序. 地质科学, 2017, 52(3): 637-648.

8 郭华东, 王力哲, 陈方, 等. 科学大数据与数字地球. 科学通报, 2014, 59(12): 1047-1054.

9 龚健雅, 李小龙, 吴华意. 实时GIS时空数据模型. 测绘学报, 2014, (3): 226-232.

10 陈建平, 李婧, 崔宁, 等. 大数据背景下地质云的构建与应用. 地质通报, 2015, 34(7): 1260-1265.

11 吴冲龙, 刘刚. "玻璃地球"建设的现状、问题、趋势与对策. 地质通报, 2015, 34(7): 1280-1287.

12 周永章, 陈烁, 张旗, 等. 大数据与数学地球科学研究进展——大数据与数学地球科学专题代序. 岩石学报, 2018, 34(2): 255-263 .

13 Han J. Data Mining: Concepts and Techniques. San Francisco: Morgan Kaufmann Publishers, 2005.

14 李德仁, 张良培, 夏桂松. 遥感大数据自动分析与数据挖掘. 测绘学报, 2014, 43(12): 1211-1216.

15 周永章, 黎培兴, 王树功, 等. 矿床大数据及智能矿床模型研究背景与进展. 矿物岩石地球化学通报, 2017, 36(2): 327-331.

16 李超岭, 李健强, 张宏春, 等. 智能地质调查大数据应用体系架构与关键技术. 地质通报, 2015, 34(7): 1288-1299.

17 于萍萍, 陈建平, 柴福山, 等. 基于地质大数据理念的模型驱动矿产资源定量预测. 地质通报, 2015, 34(7): 1333-1343.

18 Han J, Tung A K H, He J. SPARC: Spatial Association Rule-Based Classification. In: Data Mining for Scientific and Engineering Applications. New York: Springer US, 2001: 461-485.

19 吴冲龙, 何珍文, 翁正平, 等. 地质数据三维可视化的属性、分类和关键技术. 地质通报, 2011, 30(5): 642-649.

20 赵节霞, 邹毅. 三维可视化技术在数字矿山中的应用研究. 矿山测量, 2015, (2): 3-5.

Big Data Epoch: Challenges and Opportunities for Geology

ZHAI Mingguo[1,2*] YANG Shufeng[1*] CHEN Ninghua[1] CHEN Hanlin[1]

(1 School of Earth Sciences, Zhejiang University, Hangzhou 310027, China;

2 Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100012, China)

Abstract Big data has influenced the way people live and changed the method to understand and research the world. As a typical data-intensive subject, geology is facing unprecedented challenges and opportunities. In response to these challenges, geologists need not only to improve the traditional methods of research, but also to convert the intrinsic thinking patterns and embrace the big data epoch. The combination of geology and big data greatly expands the cognitive space of geology and improves the ability to acquire new knowledge of geology. At the same time, it provides new vitality for the energy mineral survey, the rational use of environmental resources and the social production and public service, which are supported by geology. On the basis of analyzing the current research status of big data in geology, this paper elaborates the frontier scientific problems of big data research in geology in China, puts forward the strategic target of the development of big data of geology, and probes into the main problems and solutions to the development of the big data of geology. Big data will change the way geologists think. The data-driven scientific discovery model will bring a new look to the development of geology. This paper calls on Chinese geological community to give more attention and support to big data.

Keywords geology, big data, data-intensive, data mining

*Corresponding author



翟明国 前寒武纪与变质地质学家，岩石学家，中国科学院院士。1976年毕业于西北大学。1982年和1989年在中国科学院研究生院和中国科学院地质研究所获理学硕士学位和理学博士学位。2009年当选中国科学院院士，2014年当选发展中国家科学院院士。中国科学院大学资深讲座教授、中国科学院地质与地球物理研究所研究员、西北大学讲座教授、浙江大学讲座教授。曾任中国科学院地质与地球物理研究所副所长、中国矿物资源探查研究中心主任、中国科学院矿产资源重点实验室主任、大陆动力学国家重点实验室主任。主要从事前寒武纪地质学、变质地质学和火成岩石学研究，针对早期陆壳的形成与增长、早期大陆壳的结构与成分、前寒武纪与显生宙变质作用和大陆动力学机制的异同等前沿核心科学问题，并且研究大陆演化历史与矿产资源的富集机理，集中研究华北陆块以及相邻地区，取得了系统的创造性成果。E-mail: mgzhai@mail.iggcas.ac.cn

ZHAI Mingguo Received the M.Sc degree and the Ph.D. in geology from Chinese Academy of Sciences (CAS), Beijing, China, in1982 and 1989, respectively. He is a Research Professor at the Institute of Geology and Geophysics, Chinese Academy of Sciences; a Chair Professor at Northwest University (China) and Zhejiang University (China); Head of State Key Laboratory of Continental Dynamics (China). His main research projects in the past decades focused on the followings: Precambrian tectonic evolution in China; high-temperature and high-pressure metamorphism and reconstruction of Precambrian; ultra-high-pressure metamorphism and continental collision; Mesozoic tectonic evolution of the North China Block. E-mail: mgzhai@mail.iggcas.ac.cn



杨树锋 构造地质学家，中国科学院院士。1975年毕业于南京大学地质系获学士学位，1981年毕业于南京大学地质系获硕士学位，1984年毕业于南京大学地质系获博士学位，是我国自己培养的第一位地质学博士。2015年当选为中国科学院地学部院士。浙江大学地球科学学院教授、博士生导师。现任教育部含油气盆地构造研究中心主任，《高校地质学报》和*Frontier of Earth Science*杂志副主编、浙江省地质学会副理事长、浙江省石油学会副理事长。30多年来一直致力于花岗岩成因机制、华南大地构造演化、中国西部造山带与盆地等方面的科学研究和人才培养工作。E-mail: yjsy-ysf@sun.zju.edu.cn

YANG Shufeng Received the B.S degree, M.Sc degree, and Ph.D. in geology from Nanjing University, China, in 1975, 1982, and 1984, respectively. He is the very first doctoriate degree receiver of geology trained in China. He is a Research Professor at Zhejiang University and academician of Chinese Academy of Sciences (CAS). He is currently the director of Research Centre for Structures in Oil and Gas Bearing Basins, associate editor of *Geological Journal of China Universities* and *Frontier of Earth Science*, and the vice chairman of Zhejiang Geological Society and Zhejiang Petroleum Society. For more than 30 years, he has been committed to scientific research and talent cultivation in the genesis of granite, tectonic evolution of the South China, and orogenic belts and basins in western China. E-mail: yjsy-ysf@sun.zju.edu.cn

■责任编辑：张帆